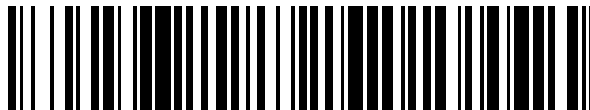


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 741 009**

51 Int. Cl.:

**G10L 19/08** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.04.2016 PCT/EP2016/057514**

87 Fecha y número de publicación internacional: **13.10.2016 WO16162375**

96 Fecha de presentación y número de la solicitud europea: **06.04.2016 E 16714448 (4)**

97 Fecha y número de publicación de la concesión europea: **15.05.2019 EP 3281197**

54 Título: **Codificador de audio y método para codificar una señal de audio**

30 Prioridad:

**09.04.2015 EP 15163055**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

**07.02.2020**

73 Titular/es:

**FRAUNHOFER GESELLSCHAFT ZUR  
FÖRDERUNG DER ANGEWANDTEN  
FORSCHUNG E.V. (100.0%)  
Hansastraße 27C  
80686 München, DE**

72 Inventor/es:

**BÄCKSTRÖM, TOM y  
JOKINEN, EMMA**

74 Agente/Representante:

**ARIZTI ACHA, Monica**

ES 2 741 009 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Codificador de audio y método para codificar una señal de audio

- 5 Realizaciones se refieren a un codificador de audio para proporcionar una representación codificada con base en una señal de audio. Realizaciones adicionales se refieren a un método para proporcionar una representación codificada con base en una señal de audio. Algunas realizaciones se refieren a una supresión de ruido de extremo lejano, de baja complejidad, de bajo retardo para códecs perceptuales de voz y audio.
- 10 Un problema actual con códecs de voz y audio es que se utilizan en entornos adversos donde la señal de entrada acústica se distorsiona por ruido de fondo y otras distorsiones. Esto provoca varios problemas. Ya que el códec ahora tiene que codificar tanto la señal deseada como las distorsiones indeseadas, el problema de codificación es más complicado debido a que la señal ahora consiste de dos fuentes y que disminuirá la calidad de codificación. Pero incluso si se puede codificar la combinación de los dos cursos con la misma calidad como una sola señal
- 15 limpia, la parte de voz aún será de menor calidad que la señal limpia. La calidad de codificación perdida no sólo es molesta de forma perceptual sino, de forma importante, también incrementa el esfuerzo de escucha, y, en el peor de los casos, disminuye la inteligibilidad o incrementa el esfuerzo de escucha de la señal decodificada.
- 20 El documento WO 2005/031709 A1 muestra un método de codificación de voz que aplica reducción de ruido al modificar la ganancia del libro de códigos. En detalle, una señal acústica que contiene un componente de voz y un componente de ruido se codifica al utilizar un análisis a través de método de síntesis, en donde para codificar la señal acústica se compara una señal sintetizada con la señal acústica durante un intervalo de tiempo, dicha señal sintetizada que se describe al utilizar un libro de códigos fijo y una ganancia fija asociada.
- 25 El documento US 2011/076968 A1 muestra un dispositivo de comunicación con codificación de voz de ruido reducido. El dispositivo de comunicación incluye una memoria, una interfaz de entrada, un módulo de procesamiento, y un transmisor. El módulo de procesamiento recibe una señal digital de la interfaz de entrada, en donde la señal digital incluye un componente de señal digital deseado y un componente de señal digital no deseado. El módulo de procesamiento identifica uno de una pluralidad de libros de códigos con base en el componente de
- 30 señal digital no deseado. El módulo de procesamiento entonces identifica una entrada del libro de códigos de la pluralidad de libros de códigos con base en el componente de señal digital deseado para producir una entrada de libro de códigos seleccionada. El módulo de procesamiento entonces genera una señal codificada con base en la entrada del libro de códigos seleccionada, en donde la señal codificada incluye una representación sustancialmente no atenuada del componente de señal digital deseado y una representación atenuada del componente de señal
- 35 digital no deseado.
- 40 El documento US 2001/001140 A1 muestra un enfoque modular para mejora de voz con una aplicación para codificación de voz. Un codificador de voz separa voz digitalizada de entrada en partes de componente en un intervalo de forma por intervalos. Las partes de componente incluyen componentes de ganancia, componentes de espectro y componentes de señal de excitación. Un conjunto de sistemas de mejora de voz dentro del codificador de voz procesan las partes de componente de tal forma que cada parte de componente tiene su propio proceso de mejora de voz individual. Por ejemplo, un proceso de mejora de voz se puede aplicar para analizar los componentes de espectro y otro proceso de mejora de voz se puede utilizar para analizar los componentes de señal de excitación.
- 45 El documento US 5.680.508 A da a conocer una mejora de codificación de voz en ruido de fondo para codificador de voz de baja velocidad. Un sistema de codificación de voz emplea mediciones de características robustas de cuadros de voz cuya distribución no se afecta fuertemente por ruido/niveles para tomar decisiones de voz para voz de entrada que se produce en un entorno ruidoso. El análisis de programación lineal de las características robustas y factores de ponderación respectivos se utilizan para determinar una combinación lineal óptima de estas
- 50 características. Los vectores de voz de entrada se hacen coincidir con un vocabulario de palabras de código a fin de seleccionar la palabra de código óptimamente coincidente, correspondiente. Se utiliza cuantificación vectorial adaptativa en la cual un vocabulario de palabras obtenido en un entorno silencioso se actualiza con base en una estimación de ruido de un entorno ruidoso en el cual se produce la voz de entrada, y entonces se busca el vocabulario "ruidoso" para la mejor coincidencia con un vector de voz de entrada. El índice de palabra de código
- 55 limpio, correspondiente entonces se selecciona para transmisión y para síntesis en el extremo receptor.
- 60 El documento US 2006/116874 A1 muestra un pos-filtrado dependiente del ruido. Un método implica proporcionar un filtro adecuado para reducción de distorsión provocada por codificación de voz, estimar ruido acústico en la señal de voz, adaptar el filtro en respuesta al ruido acústico estimado para obtener un filtro adaptado, y aplicar el filtro adaptado a la señal de voz para reducir ruido acústico y distorsión provocada por codificación de voz en la señal de voz.
- El documento US 6.385.573 B1 muestra una compensación de inclinación adaptativa para voz sintetizada residual. Un códec de voz de múltiples velocidades soporta una pluralidad de modos de velocidad de bits de codificación al

5 seleccionar adaptativamente modos de velocidad de bits de codificación para hacer coincidir restricciones de canal de comunicación. En modos de codificación de velocidad de bits más alta, una representación precisa de voz a través de CELP (predicción lineal excitada por código) y otros parámetros de modelación asociados se generan para reproducción y decodificación de mayor calidad. Para lograr alta calidad en modos de codificación de menor velocidad de bits, el codificador de voz separa la forma de onda estricta que coincide con los criterios de codificadores CELP regulares y se esfuerza para identificar características perceptuales significativas de la señal de entrada.

10 El documento US 5.845.244 A se refiere a un nivel de enmascaramiento de ruido adaptativo en análisis por síntesis que emplea ponderación perceptual. En un codificador de voz de análisis por síntesis que emplea un filtro de ponderación perceptual de corto plazo, los valores de los coeficientes de expansión espectral se adaptan dinámicamente con base en los parámetros espectrales obtenidos durante análisis de predicción lineal de corto plazo. Los parámetros espectrales que sirven en esta adaptación pueden comprender en particular parámetros representativos de la inclinación total del espectro de la señal de voz, y parámetros representativos del carácter resonante del filtro de síntesis de corto plazo.

15 El documento US 4.133.976 A muestra una codificación de señal de voz predictiva con efectos de ruido reducidos. Un procesador predictivo de señales de voz ofrece un filtro adaptativo en una red de retroalimentación alrededor del cuantificador. El filtro adaptativo combina esencialmente la señal de error de cuantificación, las señales de parámetros de predicción relacionadas formantes y la señal de diferencia para concentrar el ruido de error de cuantificación en picos espectrales que corresponden a las porciones formantes variables en el tiempo del espectro de voz por lo que el ruido de cuantificación se enmascara por los formantes de señal de voz.

20 El documento WO 9425959 A1 muestra el uso de un modelo auditivo para mejorar calidad o disminuir la velocidad de bits de sistemas de síntesis de voz. Se reemplaza un filtro de ponderación con un modelo auditivo que permite la búsqueda de vector de código estocástico óptimo en el dominio psicoacústico. Un algoritmo, que se ha nombrado PERCELP (para predicción lineal excitada de libros de códigos aleatorio, perceptivamente mejorada), se da a conocer que produce voz que es de calidad considerablemente mejor que aquella obtenida con un filtro de ponderación.

25 El documento US 2008/312916 A1 muestra un sistema de mejora de inteligibilidad de receptor, que procesa una señal de voz de entrada para generar una señal inteligente mejorada. En el dominio de la frecuencia, el espectro FFT de la voz recibida del extremo lejano se modifica según el espectro LPC del ruido de fondo de local para generar una señal inteligente mejorada. En el dominio del tiempo, la voz se modifica según los coeficientes LPC del ruido para generar una señal inteligente mejorada.

30 El documento US 2013/030800 1A muestra un procesador de inteligibilidad de voz adaptativa, que identifica de forma adaptativa y rastrea las ubicaciones de formantes, permitiendo así que se enfatizan los formantes conforme cambian. Como resultado, estos sistemas y métodos pueden mejorar inteligibilidad de extremo cercano, incluso en entornos ruidosos.

35 El documento US 2002/116182 A1 da a conocer un método para preparar una señal de voz para la codificación. El método comprende determinar si en contenido espectral de una señal de voz de entrada es representativo de una característica espectral definida (por ejemplo, una inclinación característica definida). Un componente de filtro específico de frecuencia de un filtro de ponderación se controla con base en la determinación del contenido espectral de la señal de voz o/y su ubicación en el codificador. Un componente de filtro de ponderación de núcleo del filtro de ponderación puede mantenerse independientemente del contenido espectral de la señal de voz.

40 El documento US 2009/265167 A1 da a conocer un dispositivo de codificación de audio que puede ajustar una inclinación de espectro de un ruido cuantificado sin cambiar el peso formante. El dispositivo incluye un HPF que extrae un componente de alta frecuencia de la región de frecuencia de una señal de audio de entrada, una unidad de cálculo de nivel de energía de alta frecuencia que calcula un nivel de energía del componente de alta frecuencia en una unidad de marco, un LPF que extrae un componente de baja frecuencia de la región de frecuencia de la señal de audio de entrada, una unidad de cálculo de nivel de baja energía que calcula un nivel de energía de un componente de baja frecuencia en una unidad de marco, una unidad de cálculo de coeficiente de corrección de inclinación multiplica la diferencia entre la SNR del componente de alta frecuencia y la SNR del componente de baja frecuencia ingresada desde un adicionador por una constante y añade un componente de sesgo al producto para calcular un coeficiente de corrección de inclinación. El coeficiente de corrección de inclinación se usa para ajustar la inclinación del espectro de un ruido cuantificado.

45 El documento US 2009/265167 A1 da a conocer un dispositivo de codificación de audio que puede ajustar una inclinación de espectro de un ruido cuantificado sin cambiar el peso formante. El dispositivo incluye un HPF que extrae un componente de alta frecuencia de la región de frecuencia de una señal de audio de entrada, una unidad de cálculo de nivel de energía de alta frecuencia que calcula un nivel de energía del componente de alta frecuencia en una unidad de marco, un LPF que extrae un componente de baja frecuencia de la región de frecuencia de la señal de audio de entrada, una unidad de cálculo de nivel de baja energía que calcula un nivel de energía de un componente de baja frecuencia en una unidad de marco, una unidad de cálculo de coeficiente de corrección de inclinación multiplica la diferencia entre la SNR del componente de alta frecuencia y la SNR del componente de baja frecuencia ingresada desde un adicionador por una constante y añade un componente de sesgo al producto para calcular un coeficiente de corrección de inclinación. El coeficiente de corrección de inclinación se usa para ajustar la inclinación del espectro de un ruido cuantificado.

50 En [Atal, Bishnu S., y Manfred R. Schroeder. "Predictive coding of speech signals and subjective error criteria". Acoustics, Speech and Signal Processing, IEEE Transactions on 27,3 (1979): 247-254] se describen y evalúan métodos para reducir la distorsión subjetiva en codificadores predictivos para señales de voz. Se obtiene calidad de voz mejorada: 1) por remoción eficiente de estructura de voz redundante relacionada con formante y tono antes de

la cuantificación, y 2) por enmascaramiento eficiente del ruido de cuantificador por la señal de voz.

En [Chen, Juin-Hwey y Allen Gersho. "Real-time vector APC speech coding at 4800 bps with adaptive postfiltering". Acoustics, Speech and Signal Processing, IEEE International Conference on ICASSP'87. Vol. 12, IEEE, 1987] se  
 5 presenta un codificador de voz APC vectorial mejorado (VAPC), que combina APC con cuantificación vectorial e incorpora análisis por síntesis, ponderación de ruido perceptual, y pos-filtrado adaptativo.

Es el objeto de la presente invención proporcionar un concepto para reducir un esfuerzo de escucha o mejorar una  
 10 calidad de señal o incrementar una inteligibilidad de una señal decodificada cuando la señal de entrada acústica se distorsiona por ruido de fondo y otras distorsiones.

Este objeto se resuelve por las reivindicaciones independientes.

Se abordan implementaciones ventajosas por las reivindicaciones dependientes.

15 Realizaciones proporcionan un codificador de audio para proporcionar una representación codificada con base en una señal de audio. El codificador de audio se configura para obtener una información de ruido que describe un ruido incluido en la señal de audio, en donde el codificador de audio se configura para codificar de forma adaptativa la señal de audio en dependencia de la información de ruido, de tal forma que la precisión de codificación es más  
 20 alta para partes de la señal de audio que están menos afectadas por el ruido incluido en la señal de audio que para partes de la señal de audio que están más afectadas por el ruido incluido en la señal de audio.

Según el concepto de la presente invención, el codificador de audio codifica de forma adaptativa la señal de audio  
 25 en dependencia de la información de ruido que describe el ruido incluido en la señal de audio, a fin de obtener una precisión de codificación más alta para aquellas partes de la señal de audio, que están menos afectadas por el ruido (por ejemplo, que tienen una mayor relación señal a ruido), que para partes de la señal de audio, que están más afectadas por el ruido (por ejemplo, que tienen una menor relación señal a ruido).

30 Los códecs de comunicación operan frecuentemente en entornos donde la señal deseada se daña por ruido de fondo. Realizaciones dadas a conocer en el presente documento abordan situaciones donde la señal del lado de emisor/codificador tiene ruido de fondo ya antes de la codificación.

Por ejemplo, según algunas realizaciones, al modificar la función objetiva perceptual de un códec la precisión de  
 35 codificación de aquellas porciones de la señal que tienen mayor relación señal a ruido (SNR) se puede incrementar, reteniendo de esta forma la calidad de las porciones libres de ruido de la señal. Al proteger las porciones de alta SNR de la señal, se puede mejorar una inteligibilidad de la señal transmitida y se puede disminuir el esfuerzo de escucha. En tanto que se implementan algoritmos de supresión de ruido convencionales como un bloque de pre-procesamiento al códec, el enfoque actual tiene dos ventajas distintas. Primero, por el ruido conjunto se pueden evitar efectos en tándem de supresión y codificación. Segundo, ya que el algoritmo propuesto se puede implementar  
 40 como una modificación de la función objetiva perceptual, es de muy baja complejidad computacional. Además, a menudo los códecs de comunicación estiman ruido de fondo para generadores de ruido de confort en cualquier caso, por lo cual ya está disponible una estimación de ruido en el códec y se puede utilizar (como información de ruido) sin ningún costo computacional adicional.

45 Realizaciones adicionales se refieren a un método para proporcionar una representación codificada con base en una señal de audio. El método comprende obtener una información de ruido que describe el ruido incluido en la señal de audio y codificar adaptativamente la señal de audio en dependencia de la información de ruido, de tal forma que la precisión de codificación es más alta para partes de la señal de audio que están menos afectadas por el ruido  
 50 incluido en la señal de audio que para partes de la señal de audio que están más afectadas por el ruido incluido en la señal de audio.

Realizaciones adicionales se refieren a un flujo de datos que transporta una representación codificada de una señal  
 55 de audio, en donde la representación codificada de la señal de audio codifica adaptativamente la señal de audio en dependencia de una información de ruido que describe un ruido incluido en la señal de audio, de tal forma que la precisión de codificación es más alta para partes de la señal de audio que están menos afectadas por el ruido incluido en la señal de audio que para partes de la señal de audio que están más afectadas por el ruido incluido en la señal de audio.

60 Las realizaciones de la presente invención se describen en el presente documento haciendo referencia a las dibujos adjuntos:

la figura 1 muestra un diagrama de bloques esquemático de un codificador de audio para proporcionar una representación codificada con base en una señal de audio, según una realización;

la figura 2A muestra un diagrama de bloques esquemático de un codificador de audio para proporcionar una representación codificada con base en una señal de voz, según una realización;

5 la figura 2B muestra un diagrama de bloques esquemático de un determinante de entrada del libro de códigos, según una realización;

la figura 3 muestra un diagrama de una magnitud de una estimación de ruido y un espectro reconstruido para el ruido graficado con respecto a la frecuencia.

10 la figura 4 muestra un diagrama de una magnitud de ajustes de predicción lineal para el ruido para diferentes órdenes de predicción graficados con respecto a la frecuencia.

15 la figura 5 muestra un diagrama de una magnitud de un inverso de un filtro de ponderación original y magnitudes de inversos de filtros de ponderación propuestos que tienen diferentes órdenes de predicción graficados con respecto a la frecuencia; y

la figura 6 muestra un diagrama de flujo de un método para proporcionar una representación codificada con base en una señal de audio, según una realización.

20 Elementos iguales o equivalentes o elementos con funcionalidad igual o equivalente se denotan en la siguiente descripción por números de referencia iguales o equivalentes.

25 En la siguiente descripción, se expone una pluralidad de detalles para proporcionar una explicación más completa de realizaciones de la presente invención. Sin embargo, será evidente para un experto en la técnica que las realizaciones de la presente invención se pueden practicar sin estos detalles específicos. En otros casos, se muestran dispositivos y estructuras bien conocidas en forma de diagrama de bloques en lugar de en detalle a fin de evitar complicar realizaciones de la presente invención. Además, las características de las diferentes realizaciones descritas a continuación en el presente documento se pueden combinar entre sí a menos que se indique específicamente lo contrario.

30 La figura 1 muestra un diagrama de bloques esquemático de un codificador 100 de audio para proporcionar una representación 102 codificada (o señal de audio codificada) con base en una señal 104 de audio. El codificador 100 de audio se configura para obtener una información 106 de ruido que describe un ruido incluido en la señal 104 de audio para codificar adaptativamente la señal 104 de audio en dependencia de la información 106 de ruido de tal forma que la precisión de codificación es mayor para partes de la señal 104 de audio que están menos afectadas por el ruido incluido en la señal 104 de audio que para partes de la señal de audio que están más afectadas por el ruido incluido en la señal 104 de audio.

40 Por ejemplo, el codificador 100 de audio puede comprender un estimador 110 de ruido (o determinante de ruido o analizador de ruido) y un codificador 112. El estimador 110 de ruido se puede configurar para obtener la información 106 de ruido que describe el ruido incluido en la señal 104 de audio. El codificador 112 se puede configurar para codificar adaptativamente la señal 104 de audio en dependencia de la información 106 de ruido de tal forma que la precisión de codificación es mayor para partes de la señal 104 de audio que están menos afectadas por el ruido incluido en la señal 104 de audio que para partes de la señal 104 de audio que están más afectadas por el ruido incluido en la señal 104 de audio.

50 El estimador 110 de ruido y el codificador 112 se pueden implementar por (o utilizando) un aparato de hardware tal como, por ejemplo, un circuito integrado, un arreglo de compuertas programables en el campo, un microprocesador, una ordenador programable o un circuito electrónico.

En realizaciones, el codificador 100 de audio se puede configurar para codificar de forma simultánea la señal 104 de audio y reducir el ruido en la representación 102 codificada de la señal 104 de audio (o señal de audio codificada) al codificar adaptativamente la señal 104 de audio en dependencia de la información 106 de ruido.

55 En realizaciones, el codificador 100 de audio se puede configurar para codificar la señal 104 de audio utilizando una función objetiva perceptual. La función objetiva perceptual se puede ajustar (o modificar) en dependencia de la información 106 de ruido, que codifica de esta forma adaptativamente la señal 104 de audio en dependencia en la información 106 de ruido. La información 106 de ruido, por ejemplo, puede ser una relación señal a ruido o una forma estimada del ruido incluido en la señal 104 de audio.

60 Las realizaciones de la presente invención intentan disminuir el esfuerzo de escucha o incrementar respectivamente la inteligibilidad. En el presente documento es importante indicar que las realizaciones pueden no en general proporcionar la representación más precisa posible de la señal de entrada pero intentan transmitir estas partes de la señal de tal forma que se optimiza el esfuerzo de escucha o inteligibilidad. Específicamente, las realizaciones

pueden cambiar el timbre de la señal, pero de tal forma que la señal transmitida reduce el esfuerzo de escucha o es mejor para inteligibilidad que la señal transmitida de forma precisa.

5 Según algunas realizaciones, se modifica la función objetiva perceptual del códec. En otras palabras, las realizaciones no suprimen explícitamente ruido, sino cambian el objetivo de tal forma que la precisión es mayor en partes de la señal donde es mejor la relación señal a ruido. De forma equivalente, las realizaciones disminuyen la distorsión de señal en aquellas partes donde la SNR es alta. Los oyentes humanos pueden entonces entender más fácilmente la señal. Aquellas partes de la señal que tienen baja SNR por lo tanto se transmiten con menos precisión pero, ya que contienen principalmente ruido de todas formas, no es importante codificar estas partes de forma  
10 precisa. En otras palabras, al enfocar la precisión en partes de alta SNR, las realizaciones mejoran implícitamente la SNR de las partes de voz en tanto que disminuyen la SNR de las partes de ruido.

Se puede implementar o aplicar realizaciones en cualquier códec de voz de audio, por ejemplo, en estos códecs que emplean un modo perceptual. En efecto, según algunas realizaciones la función de ponderación perceptual se  
15 puede modificar (o ajustar) con base en la característica de ruido. Por ejemplo, la envolvente espectral promedio de la señal de ruido se puede estimar y utilizar para modificar la función objetiva perceptual.

Realizaciones dadas a conocer en el presente documento son aplicables de manera preferente a códecs de voz del tipo CELP (CELP = predicción lineal excitada por código) u otros códecs en los cuales el modelo perceptual se  
20 puede expresar por un filtro de ponderación. Sin embargo se pueden utilizar realizaciones en códecs tipo TCX (TCX = excitación codificada por transformada) así como otros códecs en el dominio de la frecuencia. Además, un caso de uso preferido de las realizaciones es codificación de voz pero también se pueden emplear realizaciones de forma más general en cualquier códec de voz y audio. Ya que la ACELP (ACELP = predicción lineal excitada por código algebraico) es una aplicación convencional, se describirá en detalle más adelante la aplicación de realizaciones en  
25 ACELP. La aplicación de realizaciones en otros códecs, que incluyen códecs en el dominio de la frecuencia entonces será evidente para aquellos expertos en la técnica.

Un enfoque convencional para supresión de ruido en códecs de voz y audio es aplicarla como un bloque de pre-  
30 procesamiento separado con el propósito de retirar ruido antes de codificación. Sin embargo, al separarlo para separar lo que es hay dos desventajas principales. Primero, ya que el supresor de ruido generalmente no sólo retira ruido sino también distorsiona la señal deseada, el códec por lo tanto intentará codificar una señal distorsionada de forma precisa. El códec por lo tanto tendrá un objetivo erróneo y se pierde la eficiencia y precisión. Esto también se puede ver como un caso de problema en tándem donde bloques posteriores producen errores independientes que se suman. Mediante codificación y supresión de ruido conjunto las realizaciones evitan problemas en tándem.  
35 Segundo, ya que el supresor de ruido se implementa convencionalmente en un bloque de pre-procesamiento separado, el retardo y la complejidad computacional son altos. En contraste con eso, ya que según las realizaciones del supresor de ruido se incorpora en el códec se puede aplicar con retardo y complejidad computacional muy bajos. Esto será especialmente benéfico a dispositivos de bajo costo que no tienen la capacidad computacional para supresión de ruido convencional.

40 La descripción analizará de forma adicional la aplicación en el contexto del códec AMR-WB (AMR-WB = banda ancha adaptativa de múltiples velocidades), debido a que es a la fecha de redacción el códec de voz más comúnmente utilizado. Las realizaciones se pueden aplicar fácilmente en la parte superior de otros códecs de voz también, tal como servicios de voz mejorados 3GPP o G.718. Se señala que un uso preferido de realizaciones es un  
45 complemento a normas existentes ya que las realizaciones se pueden aplicar a códecs sin cambiar el formato de flujo de bits.

La figura 2A muestra un diagrama de bloques esquemático de un codificador 100 de audio para proporcionar una  
50 representación 102 codificada con base en la señal 104 de voz, según una realización. El codificador 100 de audio se puede configurar para derivar una señal 120 residual de la señal 104 de voz y para codificar la señal 120 residual utilizando un libro 122 de códigos. En detalle, el codificador 100 de audio se puede configurar para seleccionar una entrada del libro de códigos de una pluralidad de entradas del libro de códigos del libro 122 de códigos para codificar la señal 120 residual en dependencia de la información 106 de ruido. Por ejemplo, el codificador 100 de audio puede comprender un determinante 124 de entrada del libro de códigos que comprende el libro 122 de códigos, donde el  
55 determinante 124 de entrada del libro de códigos se puede configurar para seleccionar una entrada del libro de códigos de una pluralidad de entradas del libro de códigos del libro 122 de códigos para codificar la señal 120 residual en dependencia de la información 106 de ruido, obteniendo así un residuo cuantificado 126.

El codificador 100 de audio se puede configurar para estimar una contribución de un tramo vocal en la señal 104 de  
60 voz y para retirar la contribución estimada del tramo vocal de la señal 104 de voz a fin de obtener la señal 120 residual. Por ejemplo, el codificador 100 de audio puede comprender un estimador 130 de tramo vocal y un removedor de tramo vocal 132. El estimador 130 de tramo vocal se puede configurar para recibir la señal 104 de voz, para estimar una contribución del tramo vocal en la señal 104 de voz y para proporcionar la contribución 128 estimada del tramo vocal en la señal 104 de voz al removedor de tramo vocal 132. El removedor de tramo vocal 132

se puede configurar para retirar la contribución 128 estimada del tramo vocal de la señal 104 de voz a fin de obtener la señal 120 residual. La contribución del tramo vocal en la señal 104 de voz se puede estimar, por ejemplo, utilizando predicción lineal.

- 5 El codificador 100 de audio se puede configurar para proporcionar el residuo 126 cuantificado y la contribución 128 estimada del tramo vocal (o parámetros de filtro que describen la contribución 128 estimada del tramo 104 vocal) como representación codificada con base en la señal de voz (o señal de voz codificada).

- 10 La figura 2B muestra un diagrama de bloques esquemático del determinante 124 de entrada del libro de códigos según una realización. El determinante 124 de entrada del libro de códigos puede comprender un optimizador 140 configurado para seleccionar la entrada del libro de códigos utilizando un filtro de ponderación perceptual W. Por ejemplo, el optimizador 140 se puede configurar para seleccionar la entrada del libro de códigos para la señal 120 residual de tal forma que se reduce (o se minimiza) un error de cuantificación ponderado, sintetizado de la señal 126 residual ponderada con el filtro de ponderación perceptual W. Por ejemplo, el optimizador 130 se puede configurar para seleccionar la entrada de libro de códigos utilizando la función de distancia:

$$\|WH(x - \hat{x})\|^2$$

- 20 en donde x representa la señal residual, en donde  $\hat{x}$  representa la señal residual cuantificada, donde W representa el filtro de ponderación perceptual, y en donde H representa un filtro de síntesis de tramo vocal cuantificado. Por lo tanto, W y H pueden ser matrices de convolución.

- 25 El determinante 124 de entrada del libro de códigos puede comprender un determinante de filtro de síntesis de tramo 144 vocal cuantificado configurado para determinar un filtro de síntesis de tramo vocal cuantificado H de la contribución estimada del tramo vocal A(z).

- 30 Además, el determinante 124 de entrada del libro de códigos puede comprender un ajustador 142 de filtro de ponderación perceptual configurado para ajustar el filtro de ponderación perceptual W de tal forma que se reduce un efecto del ruido en la selección de la entrada del libro de códigos. Por ejemplo, el filtro de ponderación perceptual W se puede ajustar de tal forma que partes de la señal de voz que están menos afectadas por el ruido se ponderan más para la selección de la entrada del libro de códigos que las partes de la señal de voz que están más afectadas por el ruido. Además, (o de forma alternativa), el filtro de ponderación perceptual W se puede ajustar de tal forma que se reduce un error entre las partes de la señal 120 residual que están menos afectadas por el ruido y las partes correspondientes de la señal del residuo 126 cuantificado.

- 35 El ajustador 142 de filtro de ponderación perceptual se puede configurar para obtener coeficientes de predicción lineal de la información (106) de ruido, para determinar de esta forma un ajuste de predicción lineal (A\_BCK), y para utilizar el ajuste de predicción lineal (A\_BCK) en el filtro de ponderación perceptual (W). Por ejemplo, el ajustador 40 142 de filtro de ponderación perceptual se puede configurar para ajustar el filtro de ponderación perceptual W utilizando la fórmula:

$$W(z) = A(z/\gamma_1)A_{BCK}(z/\gamma_2)H_{de-emf}(z)$$

- 45 en donde W representa el filtro de ponderación perceptual, en donde A representa un modelo de tramo vocal, A<sub>BCK</sub> representa el ajuste de predicción lineal, H<sub>de-emf</sub> representa un filtro de des-énfasis,  $\gamma_1 = 0,92$ , y  $\gamma_2$  es un parámetro con el cual es ajustable una cantidad de supresión de ruido. Por lo tanto, H<sub>de-emf</sub> puede ser igual a  $1/(1-0,68z^{-1})$ .

- 50 En otras palabras, el códec AMR-WB utiliza predicción lineal excitada por código algebraico (ACELP) para parametrizar la señal 104 de voz. Esto significa que la primera contribución del tramo vocal, A(z), se estima con predicción lineal y se remueve y entonces la señal residual se parametriza utilizando un libro de códigos algebraico. Para encontrar la mejor entrada del libro de códigos, se puede reducir al mínimo una distancia perceptual entre el residuo original y las entradas del libro de códigos. La función de distancia se puede describir como

$\|WH(x - \hat{x})\|^2$ , donde x y  $\hat{x}$  son los residuos original y cuantificado, W y H son las matrices de convolución que corresponden, respectivamente, a  $H(z) = 1/\hat{A}(z)$ , el filtro de síntesis de tramo vocal cuantificado y W(z), la

$$W(z) = A(z/\gamma_1)H_{de-emf}(z)$$

- 55 ponderación perceptual, que se elige convencionalmente como con  $\gamma_1 =$

0,92. El residuo  $x$  se ha calculado con el filtro de análisis de tramo vocal cuantificado.

En un escenario de aplicación, puede estar presente ruido de extremo lejano aditivo en la señal de voz entrante. Por lo tanto, la señal es  $y(t) = s(t) + n(t)$ . En este caso, tanto el modelo de tramo vocal,  $A(z)$ , como el residuo original  
 5 contienen ruido. Comenzando de la simplificación de ignorar el ruido en el modelo de tramo vocal y enfocándose en el ruido en el residuo, la idea (según una realización) es guiar la ponderación perceptual de tal forma que los efectos del ruido aditivo se reducen en la selección del residuo. Mientras que normalmente se desea que el error entre el residuo original y el cuantificado se asemeje a la envolvente espectral de voz, según realizaciones se reduce el error en la región que se considera más robusta al ruido. En otras palabras, según realizaciones, los componentes de  
 10 frecuencia que están menos dañados por el ruido se cuantifican con menos error mientras que los componentes con menores magnitudes que es probable que contengan errores del ruido tienen un factor de ponderación inferior en el proceso de cuantificación.

Para tomar en cuenta el efecto del ruido en la señal deseada, primero se necesita una estimación de la señal de ruido. La estimación de ruido es un tema clásico para el cual existen muchos métodos. Algunas realizaciones proporcionan un método de baja complejidad según el cual se utiliza información que ya existe en el codificador. En un enfoque preferido, la estimación de la forma del ruido de fondo que no se almacena para la detección de actividad de voz (VAD) se puede utilizar. Esta estimación contiene el nivel del ruido de fondo en 12 bandas de frecuencia con ancho creciente. Se puede construir un espectro a partir de esta estimación al mapearlo a una escala de frecuencia lineal con interpolación entre los puntos de datos originales. Se muestra en la figura 3 un ejemplo de la estimación de fondo original y el espectro reconstruido. En detalle, la figura 3 muestra la estimación de fondo original y el espectro reconstruido para ruido de automóvil con SNR promedio de -10 dB. A partir del espectro reconstruido se calcula la autocorrelación y se utiliza para obtener los coeficientes de predicción lineal de  $p$ -ésimo orden (LP) con la recursión de Levinson-Durbin. Ejemplos de los ajustes LP obtenidos con  $p = 2...6$  se muestran en la figura 4. En  
 20 detalle, la figura 4 muestra los ajustes de predicción lineal obtenidos para el ruido de fondo con diferentes órdenes de predicción ( $p = 2...6$ ). El ruido de fondo es ruido de automóvil con SNR promedio de -10 dB.

El ajuste LP obtenido,  $A_{BCK}(z)$  se puede utilizar como parte del filtro de ponderación de tal forma que se puede calcular un nuevo filtro de ponderación para

$$W(z) = A(z/\gamma_1)A_{BCK}(z/\gamma_2)H_{de-emf}(z)$$

30 Aquí  $\gamma_2$  es un parámetro con el cual se puede ajustar la cantidad de supresión de ruido. Con  $\gamma_2 \rightarrow 0$  es pequeño el efecto, en tanto que para  $\gamma_2 \approx 1$  se puede obtener una alta supresión de ruido.

En la figura 5, se muestra un ejemplo del inverso del filtro de ponderación original así como el inverso del filtro de ponderación propuesto con diferentes órdenes de predicción. Para la figura, no se ha utilizado el filtro de desénfasis. En otras palabras, la figura 5 muestra las respuestas de frecuencia del inverso del filtro de ponderación original y el propuesto con diferentes órdenes de predicción. El ruido de fondo es ruido de automóvil con SNR promedio de -10 dB.

40 La figura 6 muestra un diagrama de flujo del método para proporcionar una representación codificada con base en una señal de audio. El método comprende un paso 202 para obtener una información de ruido que describe un ruido incluido en la señal de audio. Además, el método 200 comprende un paso 204 para codificar adaptativamente la señal de audio en dependencia de la información de ruido de tal forma que la precisión de codificación es mayor para las partes de señal de audio que están menos afectadas por el ruido incluido en la señal de audio que partes de la señal de audio que están más afectadas para el ruido incluido en la señal de audio.

Aunque se han descrito algunos aspectos en el contexto de un aparato, es claro que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a un paso del método a una característica de un paso de método. De forma análoga, los aspectos descritos en el contexto de un paso de método también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente. Algunos o todos los pasos del método se pueden ejecutar por (o utilizando) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, uno o más de los pasos de método más importantes se pueden ejecutar por un aparato de ese tipo.

50 La señal de audio codificada inventiva se puede almacenar en un medio de almacenamiento digital o se puede transmitir en un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión alámbrico tal como Internet.

60 Dependiendo de ciertos requerimientos de implementación, las realizaciones de la invención se pueden implementar



5 en hardware o en software. La implementación se puede llevar a cabo utilizando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un Blu-Ray, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria rápida, que tienen señales de control electrónicamente legibles almacenadas en la misma, que cooperan (o son capaces de cooperar) con un sistema de ordenador programable de tal forma que se lleve a cabo el método respectivo. Por lo tanto, el medio de almacenamiento digital puede ser legible por ordenador.

10 Algunas realizaciones según la invención comprenden un portador de datos que tiene señales de control electrónicamente legibles, que son capaces de cooperar con un sistema de ordenador programable, de tal forma que se lleve a cabo uno de los métodos descritos en el presente documento.

15 En general, se pueden implementar las realizaciones de la presente invención como un producto de programa de ordenador con un código de programa, el código de programa que es operativo para llevar a cabo uno de los métodos cuando el producto de programa de ordenador se ejecuta en un ordenador. El código de programa, por ejemplo, se puede almacenar en un portador legible por máquina.

Otras realizaciones comprenden el programa de ordenador para llevar a cabo uno de los métodos descritos en el presente documento, almacenado en un portador legible por máquina.

20 En otras palabras, una realización del método inventivo, por lo tanto, es un programa de ordenador que tiene un código de programa para llevar a cabo uno de los métodos descritos en el presente documento, cuando el programa de ordenador se ejecuta en un ordenador.

25 Una realización adicional de los métodos inventivos es, por lo tanto, un portador de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa de ordenador para llevar a cabo uno de los métodos descritos en el presente documento. El portador de datos, el medio de almacenamiento digital o el medio grabado son convencionalmente tangibles y/o no transitorios.

30 Una realización adicional del método inventivo es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa de ordenador para llevar a cabo uno de los métodos descritos en el presente documento. El flujo de datos o la secuencia de señales se pueden configurar, por ejemplo para transferirse a través de una conexión de comunicación de datos, por ejemplo a través de Internet.

35 Una realización adicional comprende medio de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado o adaptado para llevar a cabo uno de los métodos descritos en el presente documento.

Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa de ordenador para llevar a cabo uno de los métodos descritos en el presente documento.

40 Una realización adicional según la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, de forma electrónica u óptica) un programa de ordenador para llevar a cabo uno de los métodos descritos en el presente documento a un receptor. El receptor, por ejemplo, puede ser un ordenador, un dispositivo móvil, un dispositivo de memoria o similares. El aparato o sistema, por ejemplo, puede comprender un servidor de archivos para transferir el programa de ordenador al receptor.

45 En algunas realizaciones, un dispositivo lógico programable (por ejemplo, un arreglo de compuertas programable en el campo) se puede utilizar para llevar a cabo algunas o todas las funcionalidades de los métodos descritos en el presente documento. En algunas realizaciones, un arreglo de compuertas programable en el campo puede cooperar con un microprocesador a fin de llevar a cabo uno de los métodos descritos en el presente documento. En general, los métodos se llevan a cabo de manera preferente por cualquier aparato de hardware.

50 El aparato descrito en el presente documento se puede implementar utilizando un aparato de hardware, o utilizando un ordenador, o utilizando una combinación de un aparato de hardware y un ordenador.

55 Los métodos descritos en el presente documento se pueden llevar a cabo utilizando un aparato de hardware, o utilizando un ordenador, o utilizando una combinación de un aparato de hardware y un ordenador.

60 Las realizaciones descritas anteriormente son simplemente ilustrativas para los principios de la presente invención. Se entiende que serán evidentes modificaciones y variaciones de los arreglos y los detalles descritos en el presente documento para aquellos expertos en la técnica. Se propone que, por lo tanto, se limite solamente por el alcance de las reivindicaciones de patente inminentes y no por los detalles específicos presentados a manera de descripción y explicación de las realizaciones en el presente documento.

**REIVINDICACIONES**

1. Codificador (100) de audio para proporcionar una representación (102) codificada con base en una señal (104) de audio, en el que el codificador (100) de audio se configura para obtener una información (106) de ruido que describe un ruido incluido en la señal (104) de audio, y en el que el codificador (100) de audio se configura para codificar adaptativamente la señal (104) de audio en dependencia de la información (106) de ruido, de tal forma que la precisión de codificación es mayor para partes de la señal (104) de audio que están menos afectadas por el ruido incluido en la señal (104) de audio que para partes de la señal (104) de audio que están más afectadas por el ruido incluido en la señal (104) de audio;
 

5

10

en el que la señal (104) de audio es una señal de voz, y en el que el codificador (100) de audio se configura para derivar una señal (120) residual de la señal (104) de voz y para codificar la señal (120) residual utilizando un libro (122) de códigos;

15

en el que el codificador (100) de audio se configura para seleccionar una entrada del libro de códigos de una pluralidad de entradas de un libro (122) de códigos para codificar la señal (120) residual en dependencia de la información (106) de ruido;

20

en el que el codificador (100) de audio se configura para seleccionar la entrada del libro de códigos utilizando un filtro de ponderación perceptual (W);

25

en el que el codificador (100) de audio se configura para ajustar el filtro de ponderación perceptual (W) de tal forma que partes de la señal (104) de voz que están menos afectadas por el ruido se ponderan más para la selección de la entrada del libro de códigos que partes de la señal (104) de voz que están más afectadas por el ruido;

30

en el que el codificador (100) de audio se configura para seleccionar la entrada del libro de códigos para la señal (120) residual de tal forma que se reduce o se minimiza un error de cuantificación ponderado, sintetizado de la señal (126) residual ponderada con el filtro de ponderación perceptual W.
2. Codificador (100) de audio según la reivindicación 1, en el que el codificador (100) de audio se configura para codificar adaptativamente la señal (104) de audio al ajustar una función objetiva perceptual utilizada para codificar la señal (104) de audio en dependencia de la información (106) de ruido.
3. Codificador (100) de audio según la reivindicación 1, en el que el codificador (100) de audio se configura para codificar de forma simultánea la señal (104) de audio y reducir el ruido en la representación (102) codificada de la señal (104) de audio, al codificar adaptativamente la señal (104) de audio en dependencia de la información (106) de ruido.
4. Codificador (100) de audio según una de las reivindicaciones 1 a 3, en el que la información (106) de ruido es una relación señal a ruido.
5. Codificador (100) de audio según una de las reivindicaciones 1 a 3, en el que la información (106) de ruido es una forma estimada del ruido incluido en la señal (104) de audio.
6. Codificador (100) de audio según una de las reivindicaciones 1 a 5, en el que el codificador (100) de audio se configura para estimar una contribución de un tramo vocal en la señal de voz, y para retirar la contribución estimada del tramo vocal de la señal (104) de voz a fin de obtener la señal (120) residual.
7. Codificador (100) de audio según la reivindicación 6, en el que el codificador (100) de audio se configura para estimar la contribución del tramo vocal en la señal (104) de voz utilizando predicción lineal.
8. Codificador (100) de audio según una de las reivindicaciones 1 a 7, en el que el codificador de audio se configura para ajustar el filtro de ponderación perceptual (W) de tal forma que se reduce un efecto de ruido en la selección de la entrada del libro de códigos.
9. Codificador (100) de audio según una de las reivindicaciones 1 a 8, en el que el codificador (100) de audio se configura para ajustar el filtro de ponderación perceptual (W) de tal forma que se reduce un error entre las partes de la señal (120) residual que están menos afectadas por el ruido y las partes correspondientes de una señal (126) residual cuantificada.
10. Codificador de audio según una de las reivindicaciones 1 a 9, en el que el codificador (100) de audio se configura para seleccionar la entrada del libro de códigos para la señal (120,x) residual de tal forma que se reduce un error de cuantificación ponderado, sintetizado de la señal residual ponderada con el filtro de

ponderación perceptual (W).

11. Codificador (100) de audio según una de las reivindicaciones 1 a 10, en el que el codificador (100) de audio se configura para seleccionar la entrada de libro de códigos utilizando la función de distancia:

5

$$\|WH(x - \hat{x})\|^2$$

en donde x representa la señal residual, en donde  $\hat{x}$  representa la señal residual cuantificada, en donde W representa el filtro de ponderación perceptual, y en donde H representa un filtro de síntesis de tramo vocal cuantificado.

- 10 12. Codificador (100) de audio según una de las reivindicaciones 1 a 11, en el que el codificador de audio se configura para utilizar una estimación de una forma del ruido que está disponible en el codificador de audio para detección de actividad de voz como la información de ruido.

- 15 13. Codificador (100) de audio según una de las reivindicaciones 1 a 12, en el que el codificador (100) de audio se configura para derivar coeficientes de predicción lineal de la información (106) de ruido, para determinar de esta forma un ajuste de predicción lineal ( $A_{BCK}$ ), y para utilizar el ajuste de predicción lineal ( $A_{BCK}$ ) en el filtro de ponderación perceptual (W).

- 20 14. Codificador de audio según la reivindicación 13, en el que el codificador de audio se configura para ajustar el filtro de ponderación perceptual utilizando la fórmula:

$$W(z) = A(z/\gamma_1)A_{BCK}(z/\gamma_2)H_{de-emf}(z)$$

25 en la que W representa el filtro de ponderación perceptual, en la que A representa un modelo de tramo vocal,  $A_{BCK}$  representa el ajuste de predicción lineal,  $H_{de-emf}$  representa un filtro de síntesis de tramo vocal cuantificado,  $\gamma_1 = 0,92$ , y  $\gamma_2$  es un parámetro con el cual es ajustable una cantidad de supresión de ruido.

15. Método para proporcionar una representación codificada con base en una señal de audio, en el que el método comprende:

30 obtener una información de ruido que describe un ruido incluido en la señal de audio; y

35 codificar adaptativamente la señal de audio en dependencia de la información de ruido, de tal forma que la precisión de codificación es mayor para partes de la señal de audio que están menos afectadas por el ruido incluido en la señal de audio que partes de la señal de audio que están más afectadas por el ruido incluido en la señal de audio, en donde los componentes de frecuencia que están menos dañados por el ruido se cuantifican con menos error en tanto que los componentes que es probable que contengan errores de ruido tienen un peso inferior en el proceso de cuantificación;

40 en donde la señal (104) de audio es una señal de voz;

45 derivar una señal (120) residual de la señal (104) de voz y codificar la señal (120) residual utilizando un libro (122) de códigos;

seleccionar una entrada del libro de códigos de una pluralidad de entradas de un libro (122) de códigos para codificar la señal (120) residual en dependencia de la información (106) de ruido;

seleccionar la entrada del libro de códigos utilizando un filtro de ponderación perceptual (W);

50 ajustar el filtro de ponderación perceptual (W) de tal forma que partes de la señal (104) de voz que están menos afectadas por el ruido se ponderan más para la selección de la entrada del libro de códigos que partes de la señal (104) de voz que están más afectadas por el ruido;

55 seleccionar la entrada del libro de códigos para la señal (120) residual de tal forma que se reduce o se minimiza un error de cuantificación ponderado, sintetizado de la señal (126) residual ponderada con el filtro de ponderación perceptual W.

16. Medio de almacenamiento digital legible por ordenador que tiene almacenado en el mismo un programa de ordenador para llevar a cabo un método según la reivindicación 15.

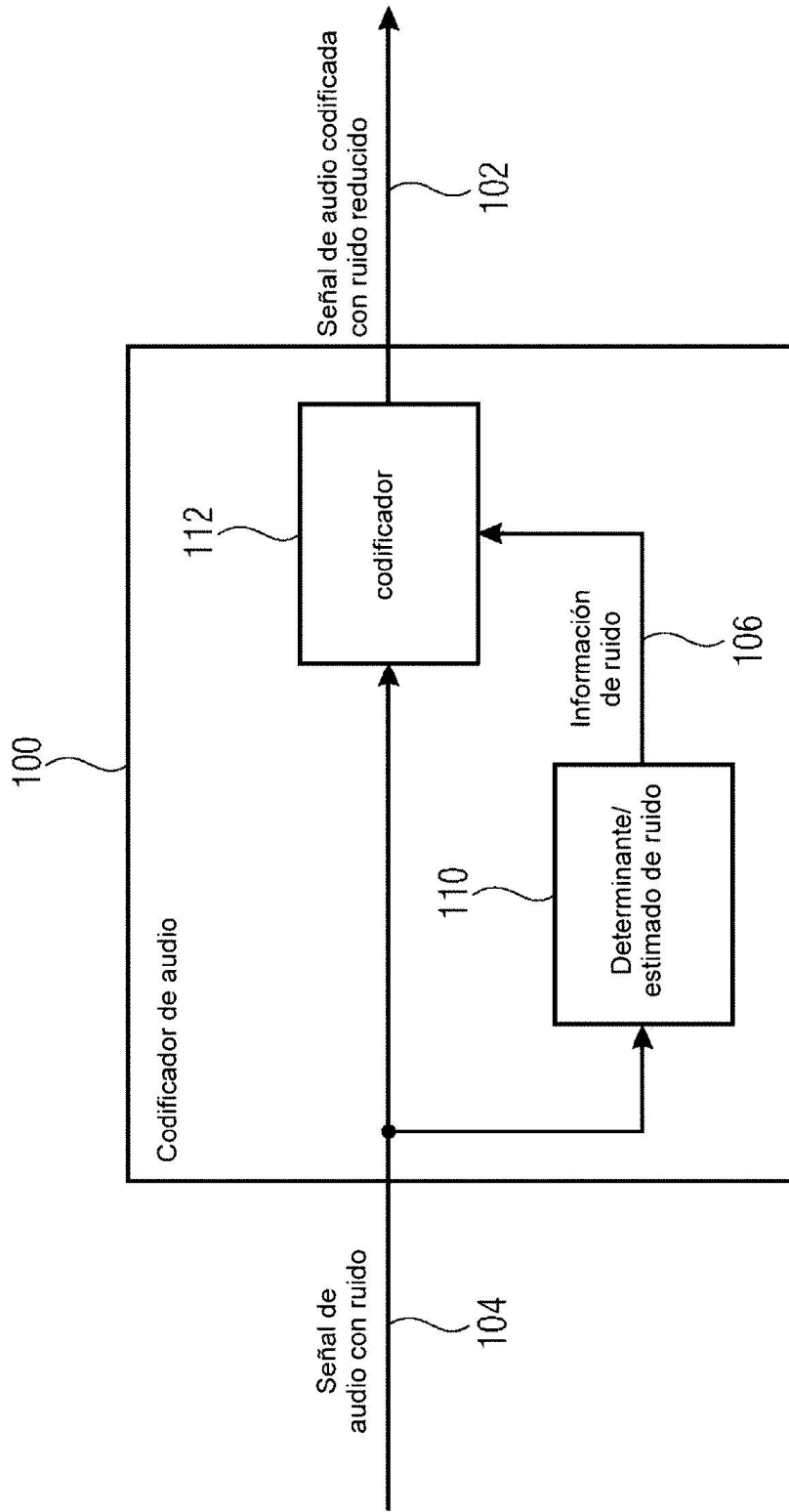


FIG 1

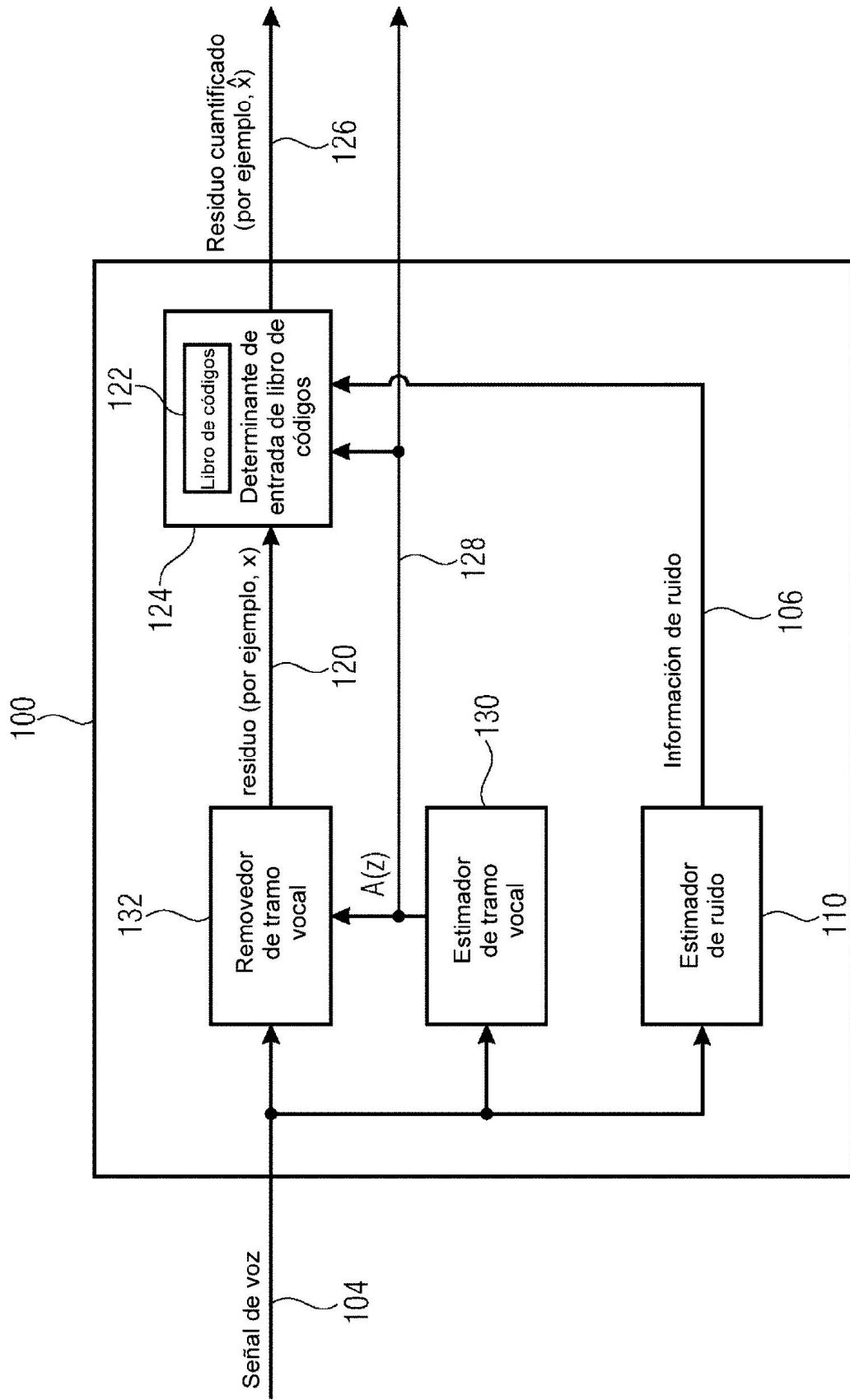


FIG 2A

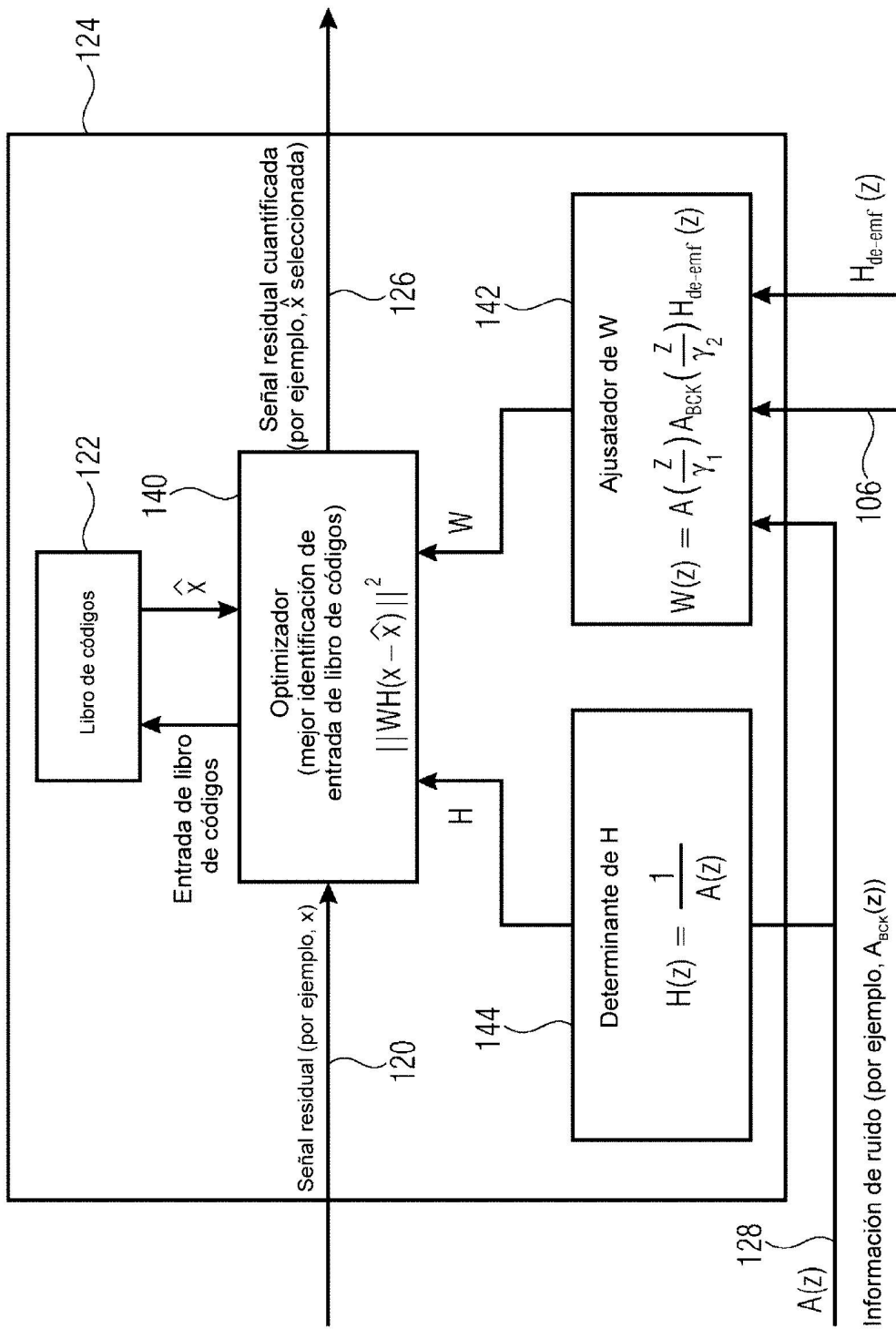


FIG 2B

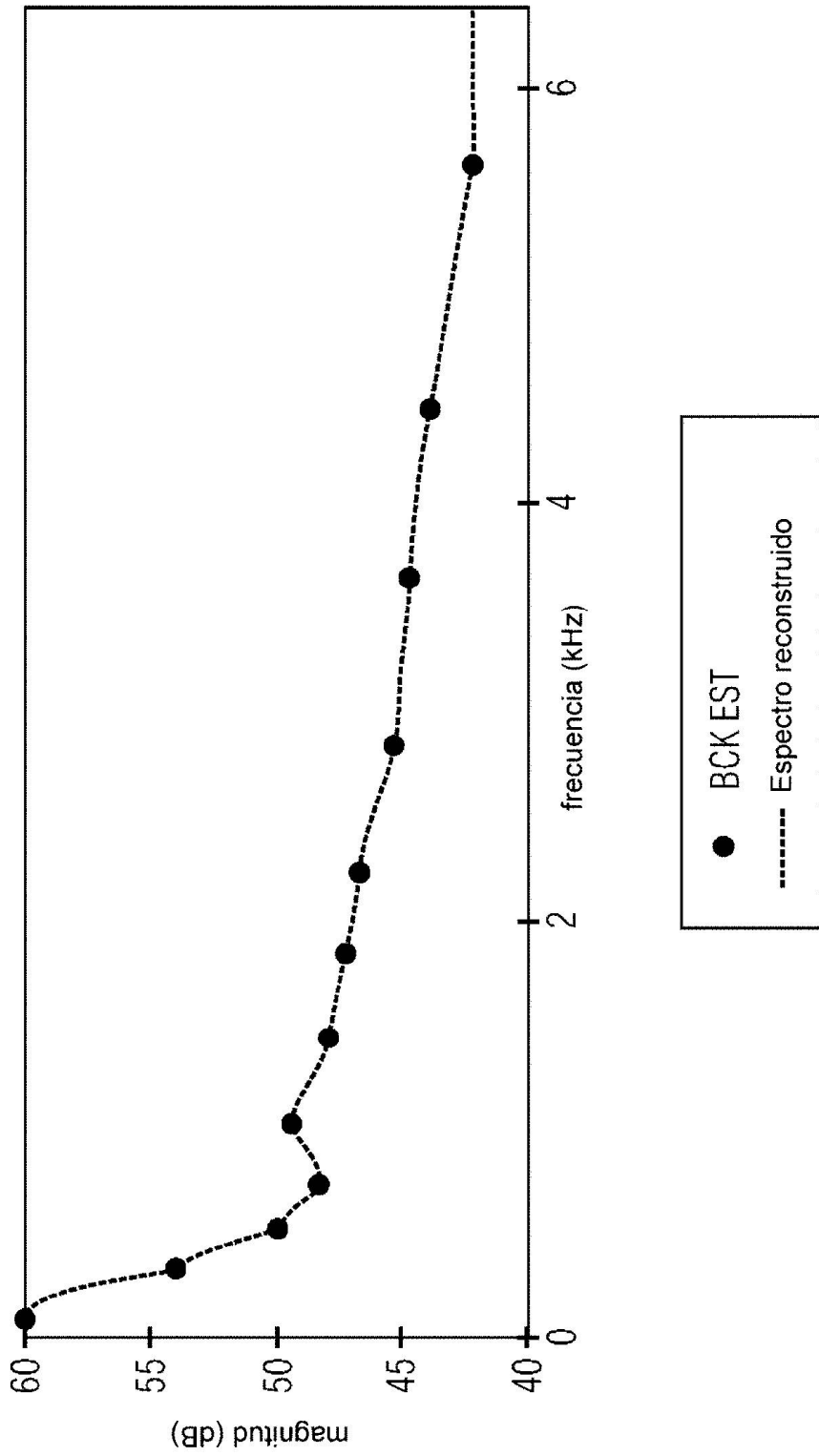


FIG 3

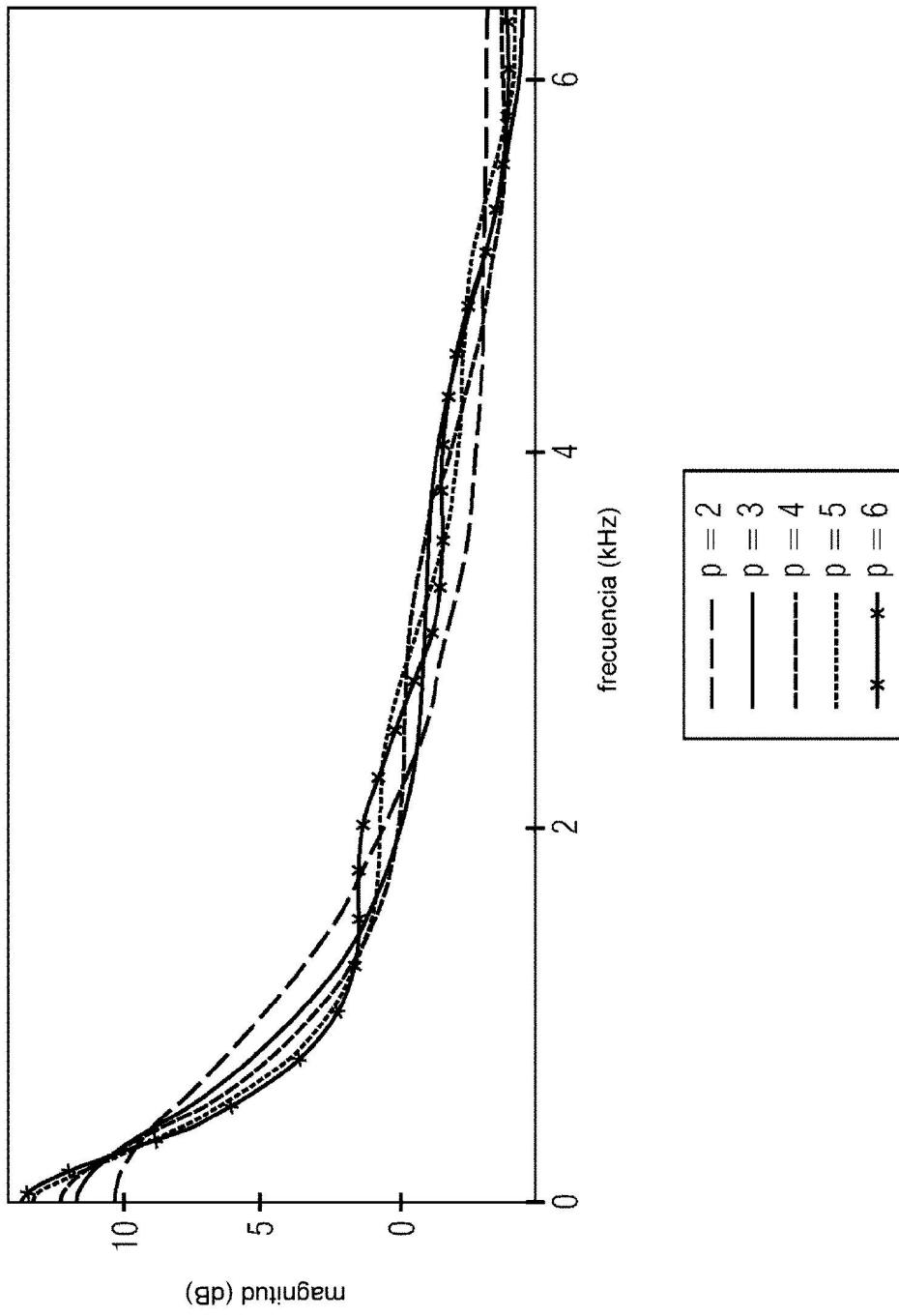


FIG 4



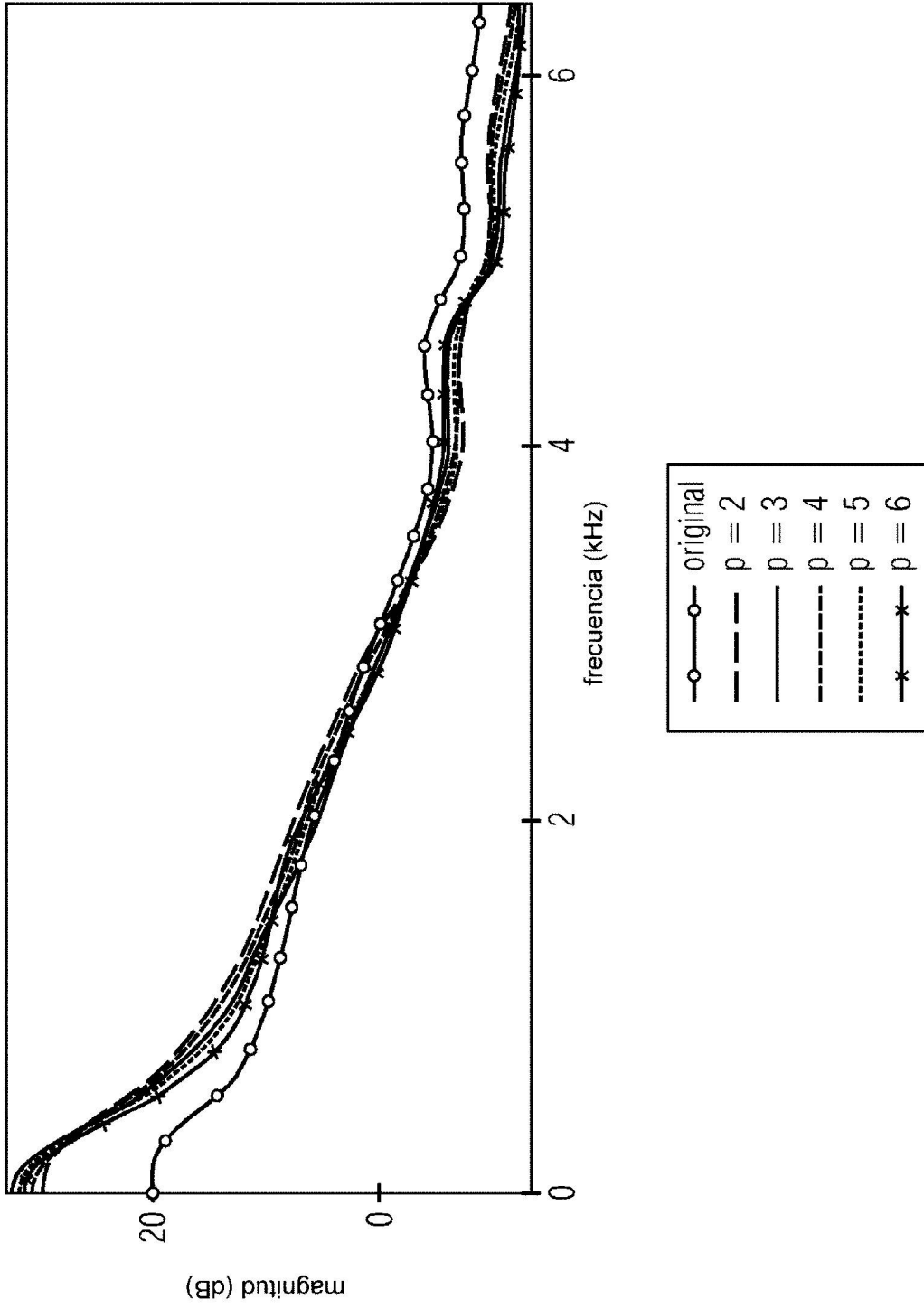


FIG 5

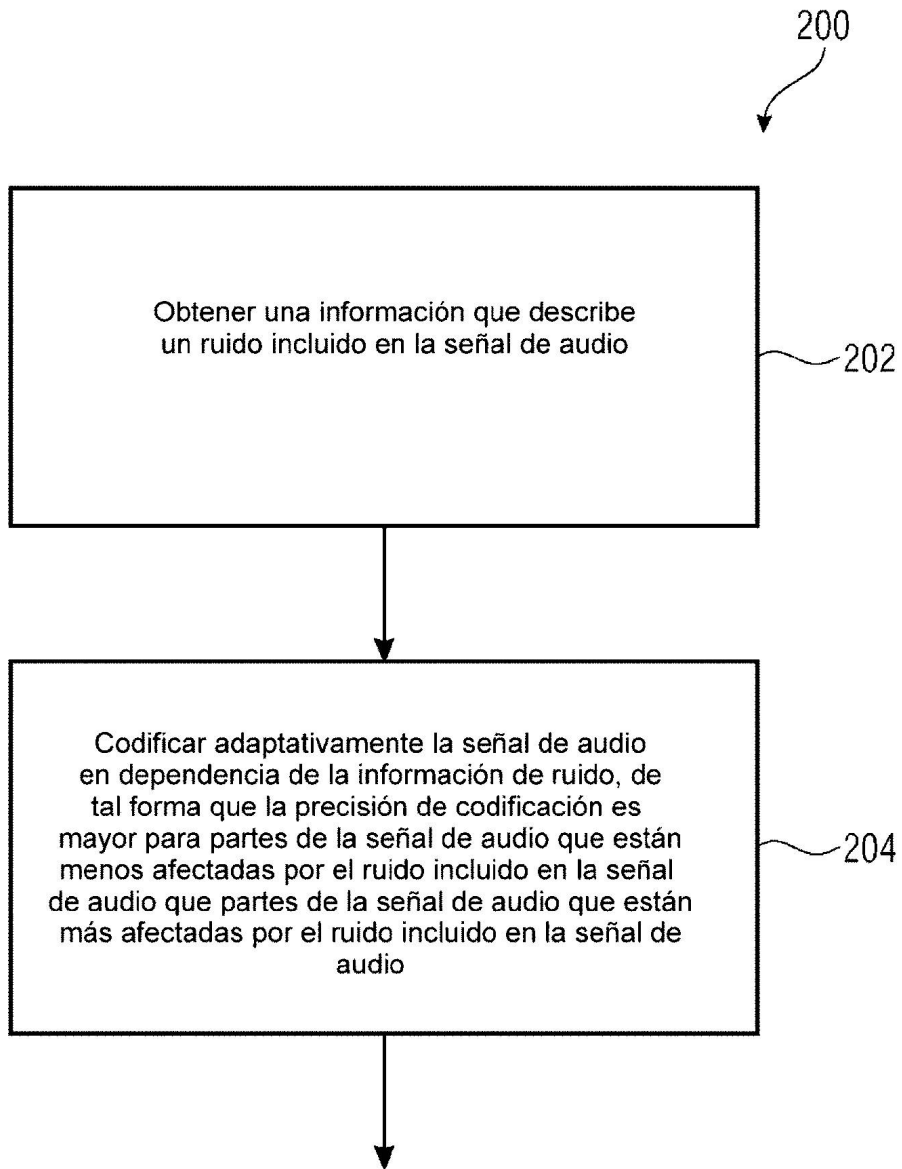


FIG 6